**JASH VAGHASIYA  
Toronto, Canada | 437-436-3011 |** [**jashmukeshvaghasiya@gmail.com**](mailto:jashmukeshvaghasiya@gmail.com) **|** [**LinkedIn**](https://linkedin.com/in/jash-vaghasiya)

**PROFILE OBJECTIVE**

Accomplished Data Engineer with 3+ years of expertise in designing and implementing enterprise-scale data solutions across Azure cloud platforms. Proven track record of optimizing data pipelines processing 850K+ daily user records, achieving up to 73% performance improvements, and reducing infrastructure costs by 35%. Specialized in real-time streaming architectures, AI/ML model integration, and data lakehouse implementations using cutting-edge technologies including Databricks, Kafka, and Spark. Microsoft Azure Data Engineer Associate and Databricks Certified Professional with demonstrated ability to translate complex data insights into strategic business value, driving data-driven decision making and identifying cost-saving opportunities.

**WORK EXPERIENCE**

**Data Engineer (August 2024 - Present)**

***Data Developer – Co-op (May 2024 - August 2024)***

[**SpicyChat.AI**](https://spicychat.ai/) **– Canada**

* Engineered and maintained enterprise-scale data pipelines processing 850K+ daily records using Azure Data Factory, PySpark on Databricks, and Azure Synapse Analytics, delivering mission-critical workflow orchestration while achieving 73% query performance improvement (45 to 12 minutes) and 69% processing time reduction (8 to 2.5 hours).
* Architected robust real-time streaming solutions leveraging Confluent Kafka and Spark Structured Streaming to capture session completion events, enabling immediate data insights and proactive performance monitoring across variable-scale data workloads.
* Spearheaded comprehensive data lakehouse implementation utilizing Unity Catalog, Azure Purview, and Delta Lake with automated compliance frameworks, ML-powered PII detection, and multi-tier validation systems that significantly reduced manual review overhead while ensuring regulatory adherence.
* Orchestrated seamless AI/ML model integration within data ecosystems, encompassing automated training data curation, validation metrics collection, and dynamic model performance dashboards in Power BI, establishing end-to-end machine learning lifecycle management.
* Designed innovative hybrid real-time/batch processing architectures that eliminated critical processing bottlenecks during peak periods while achieving 35% infrastructure cost optimization through intelligent resource allocation and automated fault-tolerance mechanisms.
* Led technical troubleshooting and production support initiatives, implementing proactive monitoring solutions and automated recovery procedures that reduced system downtime by 60% and improved SLA compliance to 99.8% while providing 24/7 support for mission-critical data systems.
* Implemented DevOps best practices through CI/CD pipeline implementation using Azure DevOps, infrastructure-as-code with Terraform, and comprehensive technical documentation, achieving 95% deployment success rates while enabling seamless knowledge transfer and reducing team onboarding time.
* Delivered actionable business intelligence by constructing scalable data ingestion pipelines from diverse sources and translating complex data insights into executive reporting frameworks, resulting in data-driven decisions that identified cost-saving opportunities.

**Data Engineer (January 2021 - November 2022)**

**Dione Apps – India**

* Developed scalable data migration strategies using Azure Data Factory, PySpark, and Synapse Analytics, enabling the seamless transfer of terabytes of data with zero data loss and minimal downtime.
* Designed and implemented end-to-end ETL pipelines integrating data from sources, ensuring data quality, business logic alignment, and high-performance transformations in Azure Databricks and SQL-based warehouses.
* Optimized data processing workflows handling 100GB+ of loads by fine-tuning PySpark transformations, reducing execution time and improving job performance and reliability.
* Implemented Slowly Changing Dimension (SCD) logic in cloud-based databases, ensuring accurate historical tracking and enhancing BI reporting reliability.
* Proactively monitored and resolved data pipeline failures using Azure Monitor, Log Analytics, and automated alerting, leading to an 80% reduction in operational issues.
* Automated DDL deployments and data ingestion workflows using CI/CD pipelines and orchestration tools, reducing manual interventions by 60% and ensuring compliance with governance standards.
* Designed custom ETL utilities to streamline data migrations between on-premise and cloud platforms, improving efficiency and standardization across projects.
* Partnered with data architects and business stakeholders to gather requirements, document processes, and ensure alignment with strategic data initiatives, driving 100% project success rates.

**CERTIFICATIONS**

* [Microsoft Certified](https://learn.microsoft.com/en-us/users/jashvaghasiya-4649/credentials/d944bfaa682ef208): Azure Data Engineer Associate (July 2024)
* [Databricks Certified](https://www.credential.net/71959e09-697b-4f64-8a15-ae7213dbf55c#gs.dwgbfp): Data Engineer Professional (August 2024)
* [Databricks Certified](https://www.credential.net/1001cd75-9e1d-4340-ba0c-78332c105021#gs.dwgbkx): Associate Developer for Apache Spark 3.0 - Python (August 2024)

**TECHNICAL SKILLS**

* **Data Engineering:** ETL/ELT Pipelines, Data Warehousing, Data Modeling (Star & Snowflake Schema), Data Governance
* **Cloud & Data Platforms:** Azure Data Factory, Power BI, Azure Synapse Analytics, Azure Data Lake, Databricks, AWS Redshift, BigQuery, Snowflake
* **Databases:** Microsoft SQL Server, PostgreSQL, MySQL, NoSQL (MongoDB, CosmosDB)
* **Programming:** Python, SQL, T-SQL, PL-SQL, PySpark, Scala, Bash, PowerShell
* **Big Data Technologies:** Apache Spark, Hadoop, Hive, Kafka
* **Workflow & DevOps:** Apache Airflow, Azure DevOps, Git, Docker, Kubernetes, Terraform
* **Monitoring & Security:** Azure Monitor, Datadog, Prometheus, RBAC, Unity Catalog, Data Encryption

**EDUCATION**

**Artificial Intelligence and Machine Learning** | Lambton College in Toronto (August 2024)

**Bachelor's Degree in Computer Applications** |Veer Narmad South Gujarat University (June 2021)